

University of Groningen

Measurement quality of the Strengths and Difficulties Questionnaire for assessing psychosocial behaviour among Dutch adolescents

Vugteveen, Jorien

DOI:
[10.33612/diss.143456742](https://doi.org/10.33612/diss.143456742)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Vugteveen, J. (2020). *Measurement quality of the Strengths and Difficulties Questionnaire for assessing psychosocial behaviour among Dutch adolescents*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.143456742>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

1

General introduction

GENERAL INTRODUCTION

Approximately 15 to 25 percent of all adolescents experience psychiatric problems, such as emotional and hyperactivity problems (Fergusson, Horwood, & Lynskey, 1993; Ormel et al., 2015). Early and accurate detection of these problems allows for timely intervention and appropriate monitoring. Detection typically occurs in one of two professional settings. The first is a community setting in which the aim is to identify adolescents at risk of psychiatric disorders among large groups of mainly healthy adolescents. This, for instance, happens during general health check-ups at schools. The second setting is a clinical setting in which the aim is to identify adolescents at risk of psychiatric disorders among adolescents with a wide range of types and severity of psychosocial problems. Moreover, in this setting the aim is to help adolescents by, amongst other things, accurately diagnosing their disorder(s) and describing the difficulties the adolescent encounters. For these purposes, healthcare professionals need information about an adolescent's psychosocial behaviour, preferably obtained from multiple informants (American Psychiatric Association, 2013). One tool that can be used in this process is the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997; Goodman, 1999).

The SDQ is one of the most widely used instruments in screening and diagnostic procedures. It was developed to measure strengths (prosocial behaviour) as well as four types of frequently occurring difficulties (emotional problems, conduct problems, hyperactivity/inattention, and social problems). Additionally, the SDQ aims to measure the impact of psychosocial problems (i.e., the chronicity, distress, social impairment for the adolescent, and burden for others) among adolescents who experience such problems. Each completed SDQ results in a total of seven scale scores: one strengths score, four difficulty scores, one total difficulties score which is the aggregate of the four difficulty scales, and one impact score. Additionally, an externalizing difficulties (the aggregate of the conduct and hyperactivity / inattention difficulties scales) and an internalizing difficulties (the aggregate of the emotional and social difficulties scales) scale score can be calculated.

Validity

An individual's SDQ scale scores, in combination with other sources of information, could give reason for action. The action can for instance pertain to referral to mental healthcare or planning diagnostic procedures. As an adolescent's mental well-being possibly depends on these actions, it is important that the interpretation of the scale scores is substantiated by evidence for their validity (Hubley & Zumbo, 2011; Messick, 1989). Validity is commonly referred to as the degree to which a test measures what it aims to measure (Kelley, 1927). Because this is a rather general description, it is useful to distinguish four types of evidence for validity (Evers et al., 1988). These types refer to the degree to which a test:

1. is subjectively regarded to cover the construct(s) it intends to measure (face validity).
2. is objectively regarded to fully cover all aspects of the construct(s) it intends to measure (content validity).
3. measures the construct(s) it aims to measure (construct validity).
4. results in scale scores that are related to relevant outcomes (criterion validity).

The combined evidence regarding these four types of validity provides an indication of the extent to which the intended interpretation of the test scores is appropriate. As the SDQ was explicitly designed with the DSM-IV (American Psychiatric Association, 1994) and ICD-10 (World Health Organization, 1992) criteria for a select number of disorders in mind (Goodman, 1997; Goodman, Meltzer, & Bailey, 1998; Goodman & Scott, 1999) and the questionnaire has been accepted by mental healthcare professionals and researchers from all around the world, I deem the SDQ's face and content validity to be sufficient. This leaves the two remaining types of validity evidence, construct and criterion validity, to be investigated.

Construct validity. Evidence for construct validity is typically gathered by assessing a standard set of three aspects (Evers, Lucassen, Meijer, & Sijsma, 2009): 1) a test's presumed internal structure, 2) whether known group differences on the construct(s) measured are indeed reflected in the group's observed test scores, and 3) a test's comparability to other tests that are supposed to measure similar constructs.

The SDQ's presumed internal structure pertains to the five scales (one for strengths and four for difficulties), each measuring one dimension of psychosocial behaviour with five items. Finding support for the SDQ's presumed internal structure would indicate that the domains measured by the five scales can be distinguished from each other, that each scale measures one domain of psychosocial behaviour, and that all items per scale contribute to measuring that domain. Evidence regarding an instrument's scale structure is typically gathered through conducting factor analysis (Asparouhov & Muthén, 2009; Evers, Sijsma, Lucassen, & Meijer, 2010; Muthén, 1984). In case of the SDQ, the analysis needs to be performed for the community and the clinical setting separately, as the validity of scale scores may be context dependent.

Additional information can be gathered by assessing the SDQ's measurement invariance across the community and clinical settings. Measurement invariance implies that SDQ scores gathered in both settings bear the same meaning and can thus be compared to each other. Comparability across settings is essential, as SDQ scores, regardless of the setting they were gathered in, are typically interpreted using community-based norm scores. Evidence regarding measurement invariance across settings can be gathered by conducting a set of multiple-group confirmatory factor analyses (Millsap & Yun-Tein, 2004).

The second construct validity aspect pertains to known groups differences. The SDQ is used in community and clinical settings among adolescents with psychiatric disorders and those without. Compared to adolescents without psychiatric disorders, it is expected that for adolescents with such disorders higher levels of difficulties and weaker levels of prosocial skills are reported. Evidence regarding this construct validity aspect can be gathered by comparing mean scale scores across these groups (Evers et al., 2010).

The third aspect involves the degree to which the SDQ compares to another test that is supposed to measure similar constructs (i.e., convergent validity evidence). This can be assessed through computing correlations between the scales of these tests (Evers et al., 2010). If the SDQ measures what it aims to measure and the other instrument does too, the scale scores of these tests should be strongly related to each other. Additional information can be obtained by comparing the SDQ to another instrument that is supposed to measure different constructs, such as intelligence or development (i.e., discriminant validity evidence). Scale scores of such an instrument and those of the SDQ should be largely unrelated to each other.

Information about the construct validity of the SDQ scales refers to the extent to which each SDQ scale score reflects the specific dimension of psychosocial behaviour that it is presumed to measure. This information helps to understand why the SDQ could be useful in screening and diagnostic procedures, because an instrument that measures what it aims to measure is more likely to be useful for its intended purposes than an instrument that does not. An instrument's value for its intended purposes is a matter of criterion validity.

Criterion validity. Criterion validity refers to the degree to which a test's scale scores are related to relevant external outcomes. The outcomes considered to gather evidence concerning this type of validity highly depend on an instrument's purpose (Evers et al., 2010). As the SDQ is used in screening and diagnostic procedures, evidence regarding its criterion validity must be gathered by investigating its value for use in these procedures. Herewith one needs to focus on identifying adolescents at risk of psychiatric disorders that are related to the domains of psychosocial behaviour covered by the SDQ: Anxiety/Mood disorder, Conduct/Oppositional Defiant Disorder (CD/ODD), Attention-Deficit/Hyperactivity Disorder (ADHD), and Autism Spectrum Disorder (ASD).

Information on how useful the SDQ is for identifying adolescents at risk of psychiatric disorders in a community setting can be obtained by investigating the SDQ's ability to distinguish between adolescents suffering from *any* of the above mentioned psychiatric disorders and adolescents that do not. Additional information can be gathered by assessing *per* disorder (Anxiety/Mood disorder, CD/ODD, ADHD, ASD) how well adolescents suffering from that disorder can be distinguished from adolescents that do not suffer from any of these disorders. The value of the separate SDQ scales for these purposes can be assessed using a receiver operating characteristic curve (Hanley &

McNeil, 1982; Metz, 1978) per scale. Jointly considering all SDQ scales for this purpose is possible using cluster analysis (e.g., Hennig, Meila, Murtagh, & Rocci, 2015), therewith investigating whether SDQ score profiles differ among the above mentioned groups of adolescents.

Compared to adolescents in a community setting, a much larger part of the adolescents in a clinical setting suffers from psychiatric disorders. Moreover, a substantial number of these adolescents likely suffer from more than one disorder, given the high comorbidity rates of psychiatric disorders among youth (Merikangas et al., 2010). Therefore, it is only marginally relevant to investigate in a clinical setting how well adolescents suffering from Anxiety/Mood disorder, CD/ODD, ADHD, or ASD can be distinguished from each other based on SDQ scores. Instead, an indication of the SDQ's value for identifying adolescents at risk of these psychiatric disorders can be obtained by assessing the extent to which SDQ scores are predictive for each of the separate disorders. Per disorder, the predictive ability of single SDQ scales can be assessed using logistic regression analysis. This approach can be extended to simultaneously including all SDQ scales and the four types of psychiatric disorders (and combinations thereof). Another way to jointly consider all SDQ scales and explicitly take into account the potential comorbidity of disorders is by using cluster analysis (e.g., Hennig et al., 2015), therewith investigating whether SDQ score profiles differ among adolescents with different (combinations of) disorders, and content-wise match the specific disorder(s) present among adolescents. For example, among adolescents suffering from both CD/ODD and ADHD a matching SDQ score profile would include high levels of conduct and hyperactivity / inattention difficulties and low scores on the three remaining scales.

Note that the above described types of evidence for the construct and criterion validity should not be viewed as separate, and possibly substitutable, pieces to the puzzle. Instead, they should be considered collectively and in relation to each other for obtaining an indication of the validity of the SDQ score interpretations.

Research aims and thesis outline

Research on validity aspects of the SDQ focusing on Dutch adolescents is scarce. As a result, little is known about how healthcare professionals in the Netherlands should interpret SDQ scores and how useful, if at all, the scores are for the SDQ's intended purposes among adolescents. In order to inform child and adolescent healthcare practice, the studies in this thesis are aimed at gathering evidence regarding construct and criterion validity aspects of the self-report and the parent-report SDQ versions for use in screening and diagnostic procedures among Dutch adolescents aged 12 to 17 years. Additionally, relative norms for interpreting scale scores of both SDQ versions are provided.

Data. This section contains a short description of the data used in the search for evidence regarding the validity aspects. This description helps understand the information provided in the final part of this paragraph, which presents the research aims of the studies described in Chapters two to six of this thesis.

Self-reported and parent-reported SDQ data were gathered in community settings and two types of clinical settings: child and adolescent social care (CASC; Dutch: Jeugdgezondheidszorg [JGZ]), and child and adolescent mental healthcare (CAMH; Dutch: Jeugd Geestelijke Gezondheidszorg [Jeugd GGZ]). Additionally, in community settings data were gathered using the Child Behavior Checklist (Achenbach, 1991a), the Youth Self Report (Achenbach, 1991b) and the Intelligence and Development Scales (Grob, Hagmann-von Arx, Ruiter, Timmerman, & Visser, 2018). Table 1.1 provides an overview of the data available per setting. Community samples 1 and 2 and CAMH sample 1 were available for all studies in this thesis. Community sample 3, CASC sample 1 and CAMH sample 2 were only available for the studies presented in Chapters 5 and 6.

Table 1.1 Available SDQ, CBCL/YSR and IDS-2 data from the community, CASC, and CAMH settings

Setting	Sample	N	SDQ			CBCL/YSR			IDS-2
			Self and parent	Only self	Only parent	Self and parent	Only self	Only parent	
Community	1	519	274	217	28	276	211	26	220
	2	443	206	220	17	192	181	1	
	3	331	292	15	24				
CASC	1	124	31	74	19				
CAMH	1	4,053	3,493	206	354				
	2	229	177	39	13				

Notes. SDQ = Strengths and Difficulties Questionnaire; CBCL = Child Behavior Checklist; YSR = Youth Self Report; IDS-2 = Intelligence and Development Scales 2; CASC = Child and adolescent social care; CAMH = Child and adolescent mental health.

Outline. Construct validity aspects are investigated starting in *Chapter 2*. Chapter 2 provides an indication of whether the SDQ scales each measure a single and distinguishable domain of psychosocial behaviour by assessing the presumed five-factor structure of the self-report and parent-report SDQ versions in community and clinical settings. Additionally, the chapter provides information on the comparability of self-reported and parent-reported SDQ scores across these settings by investigating their measurement invariance. In this study we used SDQ data collected in community (samples 1 and 2) and CAMH (sample 1) settings.

The investigation into construct validity aspects continues in *Chapter 3*. Chapter 3 focuses on using the self-report and parent-report SDQ versions in a community setting. The chapter presents further information on the SDQ versions' presumed five-scale structures when used in community settings by conducting a more in depth assessment of their factor structures. Additionally, an investigation into associations between the SDQ scales and 1) conceptually similar CBCL/YSR scales, 2) conceptually different CBCL/YSR scales, and 3) conceptually different IDS-2 scales provides an indication of the extent to which each SDQ scale measures the domain of psychosocial behaviour it is presumed to measure (i.e., convergent and discriminant validity). In this part of Chapter 3, we used SDQ, CBCL/YSR and IDS-2 data of 962 adolescents, collected in community samples 1 and 2.

Criterion validity aspects are investigated starting at the end of *Chapter 3*. That part of the chapter provides indications of how well the total difficulties scale of both informant versions can be used to distinguish between adolescents from the general population and adolescents that are at risk of psychiatric disorders. Next, the chapter provides information of how well each of the five strengths and difficulties scales of both informant versions can be used to distinguish between adolescents from the general population and adolescents diagnosed with a disorder that content-wise matches the SDQ scale (Anxiety/Mood disorder for the emotional difficulties scale, CD/ODD for the conduct difficulties scale, ADHD and the hyperactivity/inattention scale, and ASD for the social problems and prosocial behaviour scales). For this investigation we used SDQ data collected in community (samples 1 and 2) and CAMH (sample 1) settings.

The investigation into criterion validity aspects continues in *Chapter 4*. Chapter 4 focuses on using the SDQ versions in a diagnostic context by examining how well diagnosed disorders (Anxiety/Mood disorder, CD/ODD, ADHD, and ASD) can each be predicted from separate SDQ scales of both informant versions. This examination provides information on how well SDQ scales can be used to provide a preliminary indication of the type of disorder an adolescent is suffering from. For this examination, SDQ data collected in the CAMH setting (sample 1) were used.

The examinations of criterion validity aspects presented in Chapters 3 and 4 are expanded upon in *Chapter 5*. Chapter 5 focuses on using SDQ score profiles that combine all self-reported and parent-reported SDQ scales, for distinguishing between adolescents from the community, CASC and CAMH settings, and for distinguishing between diagnosed disorders, including combinations of disorders. This investigation provides an indication of how useful the SDQ score profiles are for identifying individuals at risk of psychiatric disorders in a screening context and obtaining a preliminary indication of the type of disorder(s) in a diagnostic context. In this study we used SDQ data from the community, CASC and CAMH settings (all samples).

Chapter 6 presents joint community-based relative norms and gender-specific community-based relative norms per year of age, for use among Dutch adolescents aged 12 to 17. These norms are intended for interpreting adolescent self-reported and parent-

reported SDQ scale scores gathered in community and clinical settings. The norm scores were established using SDQ data collected in the community setting (all samples).

I conclude in *Chapter 7* by discussing the main findings from the studies presented in this thesis, describing the studies' main strengths and limitations, deriving implications for practice and providing recommendations for future research.

For practical and environmental reasons, the appendices to this thesis are made available online. Links to these appendices are provided in the chapters.

